

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号  
特開2000-224590  
(P2000-224590A)

(43) 公開日 平成12年8月11日 (2000.8.11)

(51) Int.Cl.<sup>7</sup>  
H 0 4 N 7/32

識別記号

F I  
H 0 4 N 7/137

テーマコード\* (参考)

Z

審査請求 有 請求項の数18 O L (全 17 頁)

(21) 出願番号 特願平11-361271

(22) 出願日 平成11年12月20日 (1999.12.20)

(31) 優先権主張番号 0 9 / 2 3 6 8 3 8

(32) 優先日 平成11年1月25日 (1999.1.25)

(33) 優先権主張国 米国 (U S)

(71) 出願人 59706/574

ミツビシ・エレクトリック・インフォメ  
ーション・テクノロジー・センター・アメリ  
カ・インコーポレイテッド  
MITSUBISHI ELECTRIC  
INFORMATION TECHNO  
LOGY CENTER AMERIC  
A, INC.アメリカ合衆国、マサチューセッツ州、ケ  
ンブリッジ、ブロードウェイ 201

(74) 代理人 10005/874

弁理士 曾我 道照 (外6名)

最終頁に続く

(54) 【発明の名称】 ビデオシーケンスの特徴抽出方法

(57) 【要約】

【課題】 M P E G - 2 および M P E G - 4 の圧縮ビデオシーケンスからの特徴抽出の方法を提供する。

【解決手段】 圧縮ビットストリームを検査することによって、ビデオシーケンスの時間的空間的圧縮の複雑さが特徴抽出について評価され、その複雑さを、ビデオシーケンスの時間的空間的特性の記述子として用いる。時間的空間的圧縮の複雑さの尺度は、マッチング基準として用いられ、また、絶対索引付け用に用いることもできる。特徴抽出は、情景変化検出技術と共に行うことができ、この組み合わせは、合理的な正確さを有し、非常に簡単であるという利点を有する。これは、圧縮した形の信号のエントロピー復号をベースにしており、計算的に高価な逆離散的コサイン変換 (D C T) を必要としないからである。

MPEG-4 圧縮ビデオシーケンス  
フレームサイズ QCIF (176×144)  
フレーム速度 30 フレーム/秒  
圧縮速度 1-1.24倍

シーケンス	オブジェクト	源での VOP サイズ	C <sub>avg</sub> <sup>obj</sup>	C <sub>avg</sub> <sup>seq</sup>	Q <sub>avg</sub>	N <sub>obj</sub>	最も複雑な シーケンスの 尺度	最も複雑な シーケンスの 尺度
ニュース	静止背景	11×9	0.2	2.28	3	89	長い	3
ニュース	ダンサーの シーケンスを 含むTVモニタ	8×6	11.43	183.5	3	18	短い	3
ニュース	ニュース画面	11×9	1.1	42.76	3	30	短い	4
ニュース	アナウンサー	3×2	0.0	0.0	3	8	長い	1
沿岸警備隊	水	11×8	5.48	41.5	3	51	長い	4
沿岸警備隊	モータース	8×6	2.00	35.5	3	23	長い	2
沿岸警備隊	小艇のポート およびスター	11×9	4.00	83.8	3	8	長い	3
沿岸警備隊	ボートの背景	11×4	2.58	51.6	3	30	中間	4
コンテナ船	水	11×8	0.85	5.48	5	58	短い	10
コンテナ船	船	8×4	1.16	46.13	8	15	長い	4
コンテナ船	小型ボート	4×1	0.6	16.4	6	3	長い	1
コンテナ船	貨物 (積まぬ)	11×9	0.12	1.79	8	80	短い	4
コンテナ船	海上背景 (海)	11×3	0.4	2.09	6	29	長い	2
コンテナ船	海	1×1	2	97	6	0	適用なし	0
skiyo	静止背景	11×8	0.22	2.83	4	89	長い	3
skiyo	両および	8×6	1.77	35.3	4	37	長い	5

様々なMPEG-4のオブジェクトについての  
提案される記述子

## 【特許請求の範囲】

【請求項1】 少なくともIフレームのデータとPフレームのデータを含み、該Pフレームのデータが時間的に隣接するIフレームのデータを基準とする残りの相違のデータおよび移動ベクトルデータを含み、前記Iフレームのデータと前記Pフレームのデータの組み合わせが少なくとも1つのオブジェクトのビデオショットを構成する画像のフレーム単位のシーケンスを表す、圧縮した形のデジタル画像を表すビデオデータから特徴を抽出する方法であって、  
 少なくともデータのそれぞれのPフレームについて移動圧縮複雑マトリクス(CCM)を構成するステップであって、該移動CCMが、それぞれが対応するフレームにおける前記少なくとも1つのオブジェクトの対応するマクロブロックについての移動ベクトルをコード化するのに必要なビット数を表す、第1の組の数値を含むステップと、  
 少なくともデータのそれぞれのPフレームについての残りの圧縮複雑マトリクス(CCM)を構成するステップであって、該残りのCCMが、それぞれが対応するフレームにおける前記少なくとも1つのオブジェクトの対応するマクロブロックについての残りの相違をコード化するのに必要なビット数を表す、第2の組の数値を含むステップと、  
 それぞれの移動CCMについて、マクロブロック当たりの移動ベクトルのコード化に費やす平均ビット数、 $C_{avg\_mv}$ を決定するステップと、  
 それぞれの残りのCCMについて、マクロブロック当たりの残りの相違のコード化に費やす平均ビット数、 $C_{avg\_res}$ を決定するステップと、  
 それぞれの前記移動CCMについて、該移動CCMについてマクロブロック当たりの移動ベクトルのコード化に費やす前記平均ビット数を、前記移動CCMにおけるそれぞれの前記値と比較して、対応する平均よりも小さいすべての前記値をゼロに低減して対応する閾値化移動CCMを作成するステップと、  
 それぞれの該閾値化移動CCMについて、それぞれの該閾値化移動CCMにおけるゼロのランレングスを決定するステップと、  
 それぞれの該ゼロのランレングスを合計して、それぞれの前記閾値化移動CCMに対応するゼロの総数、 $N_c$ を決定するステップと、  
 それぞれの前記ゼロのランレングスを、短いランレングス、中間のランレングス、および長いランレングスの種類に分類して、それぞれの前記閾値化移動CCMについて、それぞれの前記種類におけるメンバーの数、 $N_{sr}$ 、 $N_{mr}$ 、 $N_{lr}$ を決定するステップとを含み、  
 対応するフレームの前記少なくとも1つのオブジェクトの部分のそれぞれについて、前記パラメータ $C_{avg\_mv}$ 、 $C_{avg\_res}$ 、 $N_c$ 、 $N_{sr}$ 、 $N_{mr}$ 、 $N_{lr}$ は、特徴についての

記述子を含むビデオシーケンスの特徴抽出方法。

【請求項2】 前記ステップのそれぞれをビデオショットにおけるそれぞれのPフレームに適用するステップと、  
 マクロブロック当たりの移動ベクトルのコード化に費やす平均ビット数の値を有する前記ショット内での変化の範囲が最小である、連続したフレームのグループを識別するステップと、  
 該連続したフレームのうちの1つを、前記ビデオショットの代表として選択するステップと、  
 該選択したフレームの記述子を、前記ショットについての記述子として選択するステップとを更に含む、請求項1記載のビデオシーケンスの特徴抽出方法。

【請求項3】 前記ステップのそれぞれを複数のビデオショットに適用して、前記ビデオショットのそれぞれについての記述子を作り出すステップと、  
 前記複数のショットのそれぞれについて、マクロブロック当たりの移動ベクトルのコード化に費やす平均ビット数を比較して、潜在的マッチの第1のリストを識別するステップとを更に含む、請求項2記載のビデオシーケンスの特徴抽出方法。

【請求項4】 前記ショットのそれぞれについて、すべてのランレングスの合計 $N_r$ を比較して、前記ショット間のマッチを更に識別するステップを更に含む、請求項3記載のビデオシーケンスの特徴抽出方法。

【請求項5】 前記すべてのランレングスの合計は、関係する画像におけるマクロブロックの数で割ることによって正規化する、請求項4記載のビデオシーケンスの特徴抽出方法。

【請求項6】 前記ショットのそれぞれについて最も頻繁なランレングスの特性を比較して、前記ショット間のマッチを更に識別するステップを更に含む、請求項3記載のビデオシーケンスの特徴抽出方法。

【請求項7】 前記ショットのそれぞれについてそれぞれの特定のランレングスの起こる頻度を比較して、前記ショット間のマッチを更に識別するステップを更に含む、請求項6記載のビデオシーケンスの特徴抽出方法。

【請求項8】 前記ビデオショットは、単一のビデオオブジェクトを含む、請求項2記載のビデオシーケンスの特徴抽出方法。

【請求項9】 前記ビデオショットは、それぞれのフレームにおいて多数のオブジェクトを含む、請求項2記載のビデオシーケンスの特徴抽出方法。

【請求項10】 少なくとも1フレームのデータとPフレームのデータを含み、該Pフレームのデータが時間的に隣接するIフレームのデータを基準とする残りの相違のデータおよび移動ベクトルデータを含み、前記Iフレームのデータと前記Pフレームのデータの組み合わせが少なくとも1つのオブジェクトのビデオショットを構成する画像のフレーム単位のシーケンスを表す、圧縮した

形のデジタル画像を表すビデオデータから特徴を抽出する方法であって、

少なくともデータのそれぞれのPフレームについて移動圧縮複雑マトリクス(CCM)を構成するステップであって、該移動CCMが、それぞれが対応するフレームにおける前記少なくとも1つのオブジェクトの対応するマクロブロックについての移動ベクトルをコード化するのに必要なビット数を表す、第1の組の数値を含むステップと、

それぞれの移動CCMについて、マクロブロック当たりの移動ベクトルのコード化に費やす平均ビット数、 $C_{avg\_mv}$ を決定するステップと、

それぞれの前記移動CCMについて、該移動CCMについてマクロブロック当たりの移動ベクトルのコード化に費やす前記平均ビット数を、前記移動CCMにおけるそれぞれの前記値と比較して、対応する平均よりも小さいすべての前記値をゼロに低減して対応する閾値化移動CCMを作成するステップと、

それぞれの該閾値化移動CCMについて、それぞれの該閾値化移動CCMにおけるゼロのランレングスを決定するステップと、

それぞれの該ゼロのランレングスを合計して、それぞれの前記閾値化移動CCMに対応するゼロの総数、 $N_z$ を決定するステップと、

それぞれの前記ゼロのランレングスを、短いランレングス、中間のランレングス、および長いランレングスの種類に分類して、それぞれの前記閾値化移動CCMについて、それぞれの前記種類におけるメンバーの数、 $N_{zr}$ 、 $N_{mr}$ 、 $N_{lr}$ を決定するステップとを含み、

対応するフレームの前記少なくとも1つのオブジェクトの部分のそれぞれについて、前記パラメータ $C_{avg\_mv}$ 、 $N_z$ 、 $N_{zr}$ 、 $N_{mr}$ 、 $N_{lr}$ は、特徴についての記述子を含むビデオシーケンスの特徴抽出方法。

【請求項11】 前記ステップのそれぞれをビデオショットにおけるそれぞれのPフレームに適用するステップと、

マクロブロック当たりの移動ベクトルのコード化に費やす平均ビット数の値を有する前記ショット内での変化の範囲が最小である、連続したフレームのグループを識別するステップと、

該連続したフレームのうちの1つを、前記ビデオショットの代表として選択するステップと、

該選択したフレームの記述子を、前記ショットについての記述子として選択するステップとを更に含む、請求項10記載のビデオシーケンスの特徴抽出方法。

【請求項12】 前記ステップのそれぞれを複数のビデオショットに適用して、該ビデオショットのそれぞれについての記述子を作り出すステップと、

前記複数のショットのそれぞれについて、マクロブロック当たりの移動ベクトルのコード化に費やす平均ビット

数を比較して、潜在的マッチの第1のリストを識別するステップとを更に含む、請求項11記載のビデオシーケンスの特徴抽出方法。

【請求項13】 前記ショットのそれぞれについて、すべてのランレングスの合計 $N_z$ を比較して、前記ショット間のマッチを更に識別するステップを更に含む、請求項12記載のビデオシーケンスの特徴抽出方法。

【請求項14】 前記すべてのランレングスの合計は、関係する画像におけるマクロブロックの数で割ることによって、正規化する、請求項13記載のビデオシーケンスの特徴抽出方法。

【請求項15】 前記ショットのそれぞれについて最も頻繁なランレングスの特性を比較して、前記ショット間のマッチを更に識別するステップを更に含む、請求項13記載のビデオシーケンスの特徴抽出方法。

【請求項16】 前記ショットのそれぞれについてそれぞれの特定のランレングスの起こる頻度を比較して、前記ショット間のマッチを更に識別するステップを更に含む、請求項15記載のビデオシーケンスの特徴抽出方法。

【請求項17】 前記ビデオショットは、単一のビデオオブジェクトを含む、請求項12記載のビデオシーケンスの特徴抽出方法。

【請求項18】 前記ビデオショットは、それぞれのフレームにおいて多数のオブジェクトを含む、請求項12記載のビデオシーケンスの特徴抽出方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、好ましくは情景変化検出と組み合わせた、例えば高精細度テレビ(HDTV: High Definition Television)放送信号において遭遇するタイプのビデオ信号シーケンス、あるいはワールドワイドウェブ通信媒体上で遭遇するもの等のその他の圧縮した形のビデオ情報用の、特徴抽出方法に関するものである。

【0002】

【従来の技術】デジタルカラービデオ信号の帯域幅を圧縮する基本的方法が、動画専門家グループ(MPEG)によって採用されている。

【0003】MPEG基準では、画像のフルフレームについての情報を時々のみ作り出すことによって、高データ圧縮度を達成している。フルの画像フレーム、または内部コード化映像は、「1フレーム」と呼ばれ、他のいかなるフレームから独立したフルフレーム情報を含む。BフレームおよびPフレームは、1フレーム同上の間でコード化され、基準アンカーフレームに関する画像の相違のみを記憶している。

【0004】通常、ビデオシーケンスのそれぞれのフレームは、より小さな画像データのブロックに分割され、それぞれのブロックは離散的コサイン変換(DCT)関

数の影響を受けて、統計的に依存した空間的ドメイン画素を独立周波数ドメインDCT係数に変換する。

【0005】それぞれの $8 \times 8$ ブロックの画素は、離散的コサイン変換(DCT)の影響を受けて、コード化信号を提供する。結果として得られる係数は通常、適応量量子化の影響を受けて、次にランレンスコード化および可変レンスコード化が行われる。従って、送信データのブロックは通常、 $8 \times 8$ のマトリクスよりも少ないコード語を含む。内部フレームコード化データ(1フレーム)のマクロブロックはまた、用いる量子化のレベル、マクロブロックのアドレスまたは場所のインジケータ、およびマクロブロックのタイプ等の情報を含む。後者の情報は、「ヘッダ」または「オーバーヘッド」情報と呼ばれている。

【0006】PまたはB内部フレームコード化に従ってコード化されるデータのブロックもまた、離散的コサイン係数のマトリクスからなる。しかしこの場合には、係数は、予測した $8 \times 8$ の画素のマトリクスと実際の $8 \times 8$ の画素のマトリクスとの間の残りまたは相違を表す。これらの係数もまた、量子化、およびランレンスコード化および可変レンスコード化の影響を受ける。フレームシーケンスにおいて、IおよびPフレームは、指定アンカーフレームである。それぞれのPフレームは、最後に現れたアンカーフレームから予測される。それぞれのBフレームは、それらの間にそのBフレームが配置されているアンカーフレームのうちの、一方または両方から予測される。予測コード化プロセスには、アンカーフレームのどのブロックが現在コード化されている予測フレームのブロックと最も密接にマッチしているかを示す、変位ベクトルの生成を含む。アンカーフレームにおけるマッチするブロックの画素データが、画素毎のベースで、コード化されているフレームのブロックから減じられ、残りが現れる。変換された残りおよびベクトルは、予測フレームについてのコード化データを含む。内部フレームコード化フレームと同様に、マクロブロックは、量子化、アドレス、およびタイプの情報を含む。

【0007】結果は、通常エネルギーが集中しており、1つのブロックにおいて少数の係数のみが、映像情報の主な部分を含む。係数は公知の方法で量子化されて、係数のダイナミックレンジを効果的に限定し、次に結果が、送信媒体に適用できるように、ランレンスコード化および可変レンスコード化される。

【0008】いわゆるMPEG-4フォーマットは、1996年11月のISO/IEC JTC1/SC29/WG11 MPEG-96/N1469の指定番号の下で、MPEG-4 ビデオVM編集に関する特別グループによってメンバーに配布された、「MPEG-4 ビデオ検証モデルバージョン5.0」において説明されている。MPEG-4 ビデオコード化フォーマットは、エンコーダにおいてフレームからフレームまでの(従来技

術の体系の場合と同様)可変ビットレートストリームを作成する。可変ビットレートストリームが固定レートチャンネルで送信されるので、ビットストリームを滑らかにするのにチャンネルバッファを用いる。バッファがオーバーフローまたはアンダーフローするのを防止するために、コード化プロセスの速度制御が用いられる。

【0009】ワールドワイドウェブ上でのビデオ配布等の新しいデジタルビデオサービスの到来と共に、ビデオシーケンスの特徴に関する情報を識別し抽出する信号処理技術がますます必要になっている。突然であれ、徐々にであれ、情景変化を識別することは、画像変化の索引付けのために有用であり、その後、情景を自動的に分析してある特定のマテリアルの特徴または特性を判定することができる。

【0010】

【発明が解決しようとする課題】将来きつと、かなりの量のデジタルビデオマテリアルが、上述のように圧縮またはコード化されたデータの形で提供されると思われる。可能な場合に、伸張したすなわち復号化した形ではなく圧縮した形でのビデオシーケンス情報上で動作すると、データサイズが小さいために、通常より迅速に処理することができる。他の処理を行う前にフルフレームで伸張しなければならないのではなく、直接圧縮データ上で動作することができる方法および技術を開発することは、有利である。

【0011】ブロック(マクロブロック)がオブジェクトの縁の境界線を含んでいる場合には、DCT係数で表される変換後のそのブロックにおけるエネルギーは、比較的大きなDC係数(マトリクスの上左隅)およびマトリクス全体にわたってランダムに分配されたAC係数を含む、ということもまた知られている。他方で、縁のないブロックは通常、同様の大きなDC係数(上左隅)および少数(例えば2つ)の隣接する、そのブロックに関連する他の係数よりも実質的に大きい、AC係数、を特徴とする。この情報は、空間的ドメインにおける画像変化に関係し、連続するフレームを比較することから得られる画像相違情報(すなわち時間的相違)と組み合わせると、1つのビデオオブジェクト(VO)を別のものから識別する要素が入り可能になる。画像のマクロブロックのDC値を用いると、その結果、オリジナルの内容の多くを保持する、オリジナルの画像をばんやりさせたバージョンになる。

【0012】従って、圧縮ビデオから索引付けする特徴抽出における従来の作業は、概してDC係数抽出を強調していた。論文“Rapid Scene Analysis on Compressed Video”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 5, No. 6, December 1995, page 533-544において、YeoおよびLiuは、全(圧縮していない)画像データのシーケンスをベースにした情景変化を検出するこれまでの努力や、他者の様々な圧縮ビデオ処

理技術を検討すると共に、MPEG-2圧縮ビデオドメインにおける情景変化検出へのアプローチを説明している。YeoおよびLiuは、オリジナルの画像の空間的に低減したバージョン、いわゆるDC画像、および圧縮ビデオから抽出したDCシーケンスを用いて、情景分析動作を促進する、ということを紹介した。彼らの「DC画像」は、オリジナルの画像の1ブロックにおける画素の平均値である画素からできており、DCシーケンスは、DC画像の結果として得られる数の低減した画素の組み合わせである。

【0013】Proc.SPIE Conf.on Storage and Retrieval for Image and Video Databases, January 1998において発表された論文において、Won他は、DC係数に費やすビットを利用してフレームにおいて縁を配置することによって、圧縮MPEG-2ビデオから特徴を抽出する方法を説明している。しかし、彼らの研究は、1フレームのみに限定されている。Kobla他は、同じ会報において、Yeo他のDC画像抽出を用いてビデオクリップを特徴づけるビデオ軌跡を形成する方法を説明している。Feng他(IEEE International Conference on Image Processing, Vol. II, pp. 821-824, Sept. 16-19, 1996)は、DC画像を抽出することなく、MPEG-2フレームのマクロブロックを横切るビット割り当てを用いて、突然の情景変化を検出する。Feng他の技術は、圧縮ビットストリームをパースするのに必要な分に加えてかなりの計算を行う必要がないので、計算的に最も簡単である。

【0014】最近出願した所有者が共通の出願において説明されている、本発明の発明者および協力者の発明によれば、Feng他のアプローチのある面とYeo他のアプローチのある面との組み合わせを用いて正確かつ簡単な情景変化検出を行う、計算的に簡単な方法が考案されている。都合よく、本発明の方法に従ってビット割り当て情報を利用する技術が、好ましくは情景変化検出技術に従って用いられて、特徴情報を抽出する。

【0015】1フレームからDC値を抽出するのは比較的簡単であるので、1フレームについてはDC画像抽出をベースにした技術がふさわしい、ということに注意するべきである。しかし、Pフレームについては、更なる計算が必要である。

【0016】DC画像抽出をベースにした技術を用いることによって、いったん情景/オブジェクト変化があると疑われるものが連続したフレーム/オブジェクトのグループ内に正確に配置されると、情景情報があると疑われるものの近傍のPフレーム情報に適切なビット割り当てをベースにした技術および/または適切なDC残り係数処理技術を適用することによって、切点を迅速かつ正確に配置する、ということが確認されている。この組み合わせの方法は、MPEG-2のシーケンスあるいはMPEG-4の多オブジェクトシーケンスのどちらにも適用することができる。MPEG-4の場合には、それ

ぞれのオブジェクトの領域を重み付け要素として用いて、フレームのそれぞれのオブジェクトにおける変化の重み付け合計を用いるのが有利である、ということがわかっている。

【0017】

【課題を解決するための手段】本方法の1態様によれば、そして好ましくは上に概要を述べた状況変化の識別に従って、フレームを横切る移動についてのビットおよび残りについてのビットの割り当てを用いて、MPEG-2の場合にはフレームについての、そしてMPEG-4の場合にはオブジェクトについての「記述子」を形成する。次にフレームの符号からビデオシーケンスについて記述子が形成される。ビデオシーケンスについてのこの記述子は、マッチング、分類、およびサーチに役立つ。この記述子は、比較的最近のMPEG-7の提案に関して適用業務があると思われる。

【0018】特に、少なくとも1フレームのデータとPフレームのデータを含み、Pフレームデータが時間的に隣接する1フレームデータを基準とする残りの相違のデータおよび移動ベクトルデータを含み、1フレームのデータとPフレームのデータの組み合わせが少なくとも1つのオブジェクトのビデオショットを構成する画像のフレーム単位のシーケンスを表す、圧縮した形のデジタル画像を表すビデオデータから特徴を抽出する方法は、少なくともデータのそれぞれのPフレームについて移動圧縮複雑マトリクス(CCM: compression complexity matrix)を構成する段階を含む。移動CCMは、それぞれが対応するフレームにおける少なくとも1つのオブジェクトの対応するマクロブロックについての移動ベクトルをコード化するのに必要なビット数を表す、第1の組の数値を含む。本方法は更に、少なくともデータのそれぞれのPフレームについての残りの圧縮複雑マトリクス(CCM)を構成する段階を含み、残りのCCMは、それぞれが対応するフレームにおける少なくとも1つのオブジェクトの対応するマクロブロックについての残りの相違をコード化するのに必要なビット数を表す、第2の組の数値を含む。この後に、それぞれの移動CCMについて、マクロブロック当たりの移動ベクトルのコード化に費やす平均ビット数、 $C^{avg}_{MV}$ を決定する段階が続く、次にそれぞれの残りのCCMについて、マクロブロック当たりの残りの相違のコード化に費やす平均ビット数、 $C^{avg}_{Res}$ を決定する段階が続く。その後、それぞれの移動CCMについて、その移動CCMについてマクロブロック当たりの移動ベクトルのコード化に費やす平均ビット数を、移動CCMにおけるそれぞれの値と比較して、対応する平均よりも小さいすべての値をゼロに低減して対応する閾値化移動CCMを作成する段階を含む。本方法は更に、それぞれの閾値化移動CCMについて、それぞれの閾値化移動CCMにおけるゼロのランレングスを決定する段階、およびそれぞれのゼロのランレングス

スを合計して、閾値化移動CCMのそれぞれのものに対応するゼロの総数、 $N_c$ を決定する段階を含む。その後、それぞれのゼロのランレングスを、短いランレングス、中間のランレングス、および長いランレングスの種類に分類して、それぞれの閾値化移動CCMについて、それぞれの種類におけるメンバーの数、 $N_{s1}$ 、 $N_{s2}$ 、 $N_{s3}$ を決定する。このようにして、対応するフレームの少なくとも1つのオブジェクトの部分のそれぞれについて、パラメータ $C_{avg\_v}$ 、 $C_{avg\_res}$ 、 $N_0$ 、 $N_{s1}$ 、 $N_{s2}$ 、 $N_{s3}$ が、特徴についての記述子を含む。

【0019】

【発明の実施の形態】実施の形態1。

(移動および残りのビット割り当てプロフィールからの特徴抽出) MPEG-2およびMPEG-4のビデオ「相互(inter)」圧縮は本質的に、ブロックのマッチングをベースにした移動圧縮と、その次に来る残りのDCTコード化からなる。残りは、現在のフレームと前のフレームとの間の相違の便利な尺度である。移動ベクトルによって、情景の移動特性が示される。移動ベクトルと残りのデータは、一緒になって、ビデオシーケンスの時間的空間的圧縮の複雑さを示す。更に、大きい移動ベクトルほどコード化するビットが多くなる(同じ量子化段階のサイズについて大きい残りほどそうであるように)ので、移動ベクトルおよび残りに費やされるビット数は、ビデオシーケンスの時間的空間的圧縮の複雑さを直接示す。ビット消費量は、エントロピー復号(可変レングスコードバース)のみを含み、逆DCTを含まないので、圧縮したビットストリームから容易に判定される。ビット消費量は、量子化段階のサイズと結びつけると、時間的空間的圧縮の複雑さの容易に計算される尺度となる。圧縮の複雑さの空間的分布は、ビデオシーケンスの時間的空間的特性によって決まる、ということに注意されたい。この時間的空間的複雑さの空間的分布を、ビデオシーケンスの索引付けのマッチング基準として用いることができる。

【0020】本発明の更なる態様によれば、それぞれのフレームについて、ビット割り当てをベースにした記述子が構成される。すなわち、それぞれのオブジェクト/フレームについて、それぞれ、そのオブジェクト/フレームにおけるそれぞれのマクロブロックの、移動ベクトルをコード化するのに必要なビット数、および残りをコード化するのに必要なビット数を含む、2つの「圧縮複雑マトリクス(compression complexity matrices)」が構成される。

【0021】従って、 $C_{mv} = \{R_{mv}(i, j)\}$ および $C_{res} = \{R_{res}(i, j)\}$ は、それぞれ移動ベクトルおよび残りに対応する速度マトリクスである。マトリクスQにはまた、それぞれのブロックについての量子化パラメータQPも記憶される。簡単のために、Pフレーム/オブジェクトのみを考える場合には、1つのフレーム

についてのビット割り当てをベースにした記述子は、以下の各段階に従って構成される。

【0022】1. Pフレームのマクロブロックが内部ブロックとしてコード化される場合には、その移動ベクトルビット消費量はゼロにセットされ、その残りのビット消費量は内部コーディングに費やすビット数にセットされる。このようにするのは、内部コード化は、移動補償の結果としてすべてゼロのブロックを作成し、その後、そのすべてゼロのブロックと現在コード化しているブロックとの間の相違(残り)をコード化することとして解釈することができるからである。

【0023】2. 移動ベクトルビット消費量は、量子化段階のサイズに直接影響されるわけではないが、量子化段階のサイズは、残りのビット消費量に直接影響を与える。従って、量子化段階サイズ値は、記述子の一部として含まれる。この値は、それぞれのマクロブロックについて異なっているかもしれず、その場合には、それぞれのブロック/行についての量子化値が、例えばマトリクスの形で、記述子の一部として含まれる。

【0024】3. フレーム/オブジェクトの、マクロブロック当たりの移動ベクトルに費やす平均ビット数 $C_{avg\_v}$ は、 $C_{mv}$ から計算することができる。すなわち、MおよびNが $16 \times 16$ のマクロブロックの数において測定される(例えば、QCIFの $176 \times 144$ のオブジェクトについて、M(幅) $=176$ でN(高さ) $=144$ )場合に、そのオブジェクトにおけるマクロブロック当たりの移動ベクトルに費やす平均ビット数の値は、99個のマクロブロックに関連する移動ベクトルの1つずつについてのビット数の値を合計し、その合計を99で割ることによって、計算することができる。

【0025】4. この平均は、オブジェクト/フレームの移動の複雑さの非常に有用な尺度である。これは第1のマッチング基準として用いられ、これによって、ターゲットのオブジェクト/フレームと比べて移動に集中する度合いがかなりずれるオブジェクト/フレームを考慮することが排除される。例えば、MPEGビデオ圧縮システムにおいて、この基準によって、静止オブジェクトと動いているオブジェクトとの間の識別が容易になる。

【0026】5. 同様に、それぞれのフレーム/オブジェクトについて、マクロブロック当たりの残りの平均の圧縮の複雑さ $C_{avg\_res}$ を計算する。この数もまた、オブジェクト/フレームを分類またはマッチさせるのに役立つ。

【0027】6. それぞれのオブジェクト/フレームにおける移動についての、マクロブロック当たりの費やす最大ビット数とマクロブロック当たりの費やす最小ビット数を決定する。

【0028】7. 以下のように、移動ベクトル速度マトリクスの「ランレングス(run-length)」属性を計算する。

【0029】(a) 移動ベクトル速度マトリクスのそれぞれの要素を、マクロブロック当たりの移動ベクトルに費やす平均ビット数の閾値(上記3を参照)と比較する。閾値よりも小さい移動ベクトル速度マトリクスのそれぞれの要素がゼロにセットされ、閾値と等しいまたはそれよりも大きいそれぞれの要素がそのままマトリクス内に入力される。移動ベクトル閾値マトリクス $C_{mv}^{thr}(i, j)$ が形成される。閾値は記述子の一部として供給される。

【0030】(b) 移動ベクトル閾値マトリクスにおけるゼロのランレングスが決定される。すべてのランレングスを合計することによって、ゼロの総数 $N_0$ が決定される。次にこれらのランレングスが、短い、中間、および長い3つの種類に分類される。これらの目的のために、QCIFの画像について、オブジェクト/フレームにおけるマクロブロックの総数の6%よりも少ないものを短いに、約9-10%までを中間に、中間の上限(例えば10%)よりも上を長いに規定するのが適当である、ということがわかっている。これらのパーセンテージの範囲は、他の場合にもまた適用することができるが、経験によって知識を得ると別の方法で定める方がよいことがわかるかもしれない。3つの種類のそれぞれにおける要素の数、すなわち $N_{sf}$ 、 $N_{mr}$ 、 $N_{lr}$ が決定される。これらの数は、マッチング用のランレングス特徴として用いられる。これらの3つの異なるレベルを種類として用いる理由は、同じフレームの並進したすなわち鏡面反射したバージョンが、根本的に異なる記述子を作成することを防止する、ということである。更に、このように分類することによって、計算が簡単になり、異なるサイズのオブジェクトまたはフレームを比較するのに役立つ。ランが概して短くよく知られた情景の例としては、顔および肩のシーケンスがあり、ランが概して長いフレームには、静止したまたは非常に移動が遅いオブジェクトのシーケンスや、激しいアクションが集中したフレームがある。移動が遅いオブジェクトと移動が速いオブジェクトとを組み合わせたフレームは、概して中間のランレングスを有するものとされる。

【0031】従って、記述子構成は、圧縮フォーマット、圧縮速度、フレームサイズ、フレーム速度に加えて、以下のものからなる。

【0032】 $C_{avg}^{mv}$ 、 $C_{mv}^{thr}$ 、 $N_0$ 、 $N_{sf}$ 、 $N_{mr}$ 、 $N_{lr}$ 、 $C_{avg}^{ros}$ 、 $C_{ros}$ 、 $Q$

【0033】従って、記述子は、シーケンスのそれぞれのフレームについて作り出される。

【0034】1ショットについての記述子は、以下の各段階によって決定される。

【0035】1. まず、上述した、より完全には、言及した同時係属の出願における、現在の情景変化検出技術を用いて、ビデオシーケンスを時間的にショットに分割する。

【0036】2. ショット全体にわたって、マクロブロック当たりの移動ベクトルに費やす平均ビット消費量、 $C$ と最も範囲が共通する( $C$ に最も近い)フレームのグループを見つける。そのグループから1フレームがランダムに選択され、その記述子を全ショットの記述子として用いる。

【0037】上述の記述子は、同様のショットのマッチング用に用いることができる。すなわち、マッチングにおける第1のパスとして、ショットのグループ内での潜在的マッチの第1のリストを、同様の(すなわち、ある数の範囲内で)マクロブロック当たりの移動ベクトルに費やす平均ビット消費量を有することをベースにして、作成することができる。

【0038】次に、それぞれの記述子におけるすべてのランレングスの合計 $N_0$ を第2のパスとして用いて、第1のパスからの候補のすべてを潜在的なマッチについて更にランク付けする。または、最も頻繁なランレングスの種類を検査してマッチを探す。最も頻繁なランレングスについて2つの候補ショットが同じ値を有する場合には、そのランレングスの頻度を用いて良好なマッチについてのランキングを行ってもよい。最もよく起こるランレングスの種類についてそれらのショットが同じ値を有しない場合には、その対は「劣等」なマッチの種類に入れられる。劣等なマッチの種類は、通常はそれ以上ランク付けされない。

【0039】前述の技術をよく知られたテスト画像(シーケンス)に適用することは、得られる記述子や行うことができる分析の性質を理解するのに有用である。その目的のために、「Ak i y o」のシーケンスを分析した。それぞれのマクロブロックが $16 \times 16$ アレイの画素からなり、 $11 \times 9$ アレイのマクロブロックから成り立っている「Ak i y o」のシーケンスの「背景」オブジェクト(オブジェクトNo. 0)に関連する移動ビット消費量を表すマトリクスは、以下に示すようになる。

【0040】移動ビット消費量

Ak i y o: (1秒当たり10フレーム)オブジェクトナンバー0

サイズ マクロブロックで $11 \times 9$ (マクロブロックサイズ $16 \times 16$ )(背景)

0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	2	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	2	0	0	0	7	6	0	0
0	0	4	0	0	0	0	0	0	8	0
0	2	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0

【0041】「背景」オブジェクトについて予想されるように、移動情報の複雑さは低い。マクロブロック当た

りの平均ビット数(9.9マクロブロック)は、0.31である。

【0042】この場合に閾値化後に得られる出力マトリクスは、閾値(0.31)が非常に低いので、入力マトリクスと同じになる。左上から右下に走査するオブジェクトのランレングス表現は、以下のようになる。

【0043】40 2 17 2 3 7 0 6 4  
4 6 8 2 2 20 (閾値T=0)

【0044】同様に、9×8アレイから成り立っている「Akियोの頭および肩」(オブジェクトNo. 1)についての移動ビット消費量は、以下のようになる。

【0045】オブジェクトナンバー1 サイズ9×8  
(Akियोの頭および肩)

0 0 0 10 21 15 0 0 0  
0 0 0 4 2 2 0 0 0  
0 0 0 22 16 18 4 0 0  
0 0 0 14 2 4 2 0 0  
0 0 6 4 22 22 2 5 0

3 10 0 21 0 15 6 4 0 2 0 2 6 22 0 16  
0 18 0 4 5 14  
0 2 0 4 0 2 4 6 0 4 0 22 0 22 0 2 0  
5 2 4 0 6 0 2  
0 2 0 29 0 6 1 4 4 2 0 2 0 2 0 2 1 (閾  
値T=0)

【0048】オブジェクトNo. 0およびオブジェクトNo. 1についての対応する残りのビット消費量は、以下のようになる。

【0049】残りのビット消費量 オブジェクトナンバー0

0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 13 0 0 0  
0 0 0 0 0 0 0 0 0 0 0  
0 0 0 8 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0  
0 16 0 0 0 0 0 0 0 0 0  
0 0 0 0 0 0 0 0 0 0 0

【0050】オブジェクトナンバー1

オブジェクトNo. 0

40 13 20 8 20 16 20 (閾値T=0.37)

【0053】

オブジェクトNo. 1

5 28 6 18 0 24 0 22 6 69 0 55 0 25 0  
14 5 56 0 96 0 61 0 21 5 28 1 26  
0 31 0 17 2 23 2 36 1 19 3 16 2 63 1  
14 6 18 0 47 0 51 0 25 1 (閾値T=13)

【0054】画像「ニュースシーケンスからのモニタ」もまた分析されている。ニュースシーケンスからのモニタ(1秒当たり10フレーム)サイズ6×5

0 4 6 2 2 29 6 0 0  
0 2 0 2 2 2 6 0 4  
0 0 0 0 2 2 2 2 0

【0046】マクロブロック当たりの平均ビット数は3.77であり、「背景」に関連するものよりかなり大きい数である。この場合の閾値化後のマトリクスは、以下のようになる。

0 0 0 10 21 15 0 0 0  
0 0 0 4 0 0 0 0 0  
0 0 0 22 16 18 4 0 0  
0 0 0 14 0 4 0 0 0  
0 0 6 4 22 22 0 5 0  
0 4 6 0 0 29 6 0 0  
0 0 0 0 0 0 6 0 4  
0 0 0 0 0 0 0 0 0

【0047】同様に、ランレングス表現もはるかに多くのデータを含み、以下のようになる。

0 0 0 5 0 28 0 0 0  
0 0 0 18 24 22 0 0 0  
0 0 0 69 55 25 14 0 0  
0 0 0 56 96 61 21 0 0  
0 0 0 28 9 26 31 17 0  
0 23 0 7 36 8 19 0 0  
0 16 0 5 63 5 14 0 0  
0 0 0 0 18 47 51 25 0

【0051】オブジェクトNo. 0についての残りに関連するマクロブロック当たりの平均ビットは0.37であり、オブジェクトNo. 1については13.08である。それぞれのランレングス表現は、以下のようになる。

【0052】

移動複雑マトリクスC<sub>mv</sub>

0 4 0 4 15 16  
2 26 7 33 6 20



0 4 32 0 26 16  
 0 0 2 2 26 21  
 0 0 0 2 2 0

【0055】平均ビット／マクロブロック＝8.86  
 モニタ

閾値化後のマトリクス

0 0 0 0 15 16  
 0 26 0 33 0 20  
 0 0 32 0 26 16  
 0 0 0 0 26 21  
 0 0 0 0 0 0

$N_0=20$ ;  $N_{s2}=7$ ;  $N_{x1}=1$ ;  $N_{x2}=0$

【0056】残りの複雑マトリクス $C_{res}$

0 19 0 49 169 33  
 7 82 33 49 248 32  
 0 24 26 0 76 0  
 0 0 48 36 64 9  
 0 0 0 14 20 0

平均ビット／マクロブロック＝94.36 QP＝12

【0057】上に示した最初の2つの例は、MPEG-4のオブジェクトについての記述子である。シーケンス「Akio」の背景等の低移動オブジェクトは、ニュースシーケンスにおけるモニタ等の高活動オブジェクトからの識別が非常に容易である、ということがわかる。

【0058】図2は、MPEG-2のフレームについての時間的空間的複雑さの例を示す。MPEG-2のフレームの特性は通常、MPEG-4のオブジェクトの特性ほど均質ではなく、従って、ばらつきが見られる。しかし、MPEG-2の場合であっても、本発明による記述子によって、時間的空間的に同様のフレームを簡単かつ効果的にクラスター化することができる。

【0059】より高解像度のデータほど高移動ベクトルになるので、より大きいMPEG-2のフレーム（MPEG-4の場合におけるオブジェクトと比較して）の移動の複雑さが見たところより高いことに注意されたい。従って、現在のマッチング基準を利用しながらフレームサイズに関する情報を保持することが重要である。この記述子が、移動ベクトルの部分を削除すれば、1フレームと同様にBフレームにも適用可能である、ということにも注意されたい。

【0060】MPEG-4のテストセットは、容易に利用可能なオブジェクトへの分割を有しているので、本発明による記述子は、主にMPEG-4のテストセットに適用されている。

【0061】オブジェクトは多かれ少なかれ均質な特性を有しているので、本発明の記述子を利用する比較によって、通例妥当な結果が生じる。同様に配置されたオブジェクトの、オブジェクト同士の比較を用いることによって、2つの多オブジェクトのMPEG-4シーケンスを比較している。基本移動補償情報が記述子に用いられ

ているので、結果は、残りのDCTコード化と結びつけたブロック移動補償を用いるいかなる圧縮ビデオ構文にも容易に適用可能なはずである。更に、分割されていないフレームもまた複合オブジェクトとみなすことができるので、このアプローチはまたこのようなフレームにも当てはまるはずである。その場合には、シーケンスを単一のオブジェクトとして扱うことによって、それらのシーケンスについての記述子が作り出される。このような「複合」オブジェクトの特性は均質ではないので、個々のオブジェクトの記述子とのどのような比較も、妥当な結果をもたらさそうにない。しかし、2組のフレームをベースにした記述子同士を比較すれば、有用な結果が提供されよう。

【0062】ターゲットのアプリケーションは、最小予想品質が高いマルチメディアデータベースであるので、本研究はまた、MPEG-1のビットレートにも集中している。本研究はまた、フルフレーム速度、すなわち1秒当たり25または30フレームに主に関心している。移動の複雑さの特徴は、フレーム速度の関数として変化する、ということに注意すべきである。しかし、移動ベクトルに費やすビット割り当ては、ビットレートが低い場合であっても、非常に変化することはない、ということが、経験的に確認されている。残りのビット割り当てのみが、ビットレートの変化によって非常に影響を受ける。また、速度制御方法が変化しても、移動特性が非常に影響を受けることはないが、残りの情報を非常に変化させる、ということもわかっている。これらの要因によって、記述子が、主に移動特性をベースにした特徴を強調するものになっている。しかし、残りの情報は、ある状況においては有用となり得る異なる情報を提供するので、残りの情報が作り出され保持される。例えば、2つのオブジェクトが同じ動作特性を有しているが、同じQP（量子化パラメータ）を用いると、一方のオブジェクトは残りのコード化に他方よりも多くのビットを必要とする場合には、前者は後方よりも時間的空間的により複雑である。しかし、一方のオブジェクトについての量子化段階のサイズが他方のオブジェクトについてのそれと等しいかそれよりも大きく、しかし第2のオブジェクトに費やす残りのコード化ビットが第1のオブジェクトについてのビットよりも少ない場合には、この2つのオブジェクトの相対的な時間的空間的複雑さに関して結論が引き出せない。すなわち、この点において、残りのビット割り当ては、必ずしも結論に役立つ情報を提供するわけではない。

【0063】普通知られている1組のMPEG-4のオブジェクトについての典型的な記述子の値を図1に示す。記述子の仕様は、この図1を参照することによって説明することができる。記述子を用いる段階のシーケンスは、いくつか異なった形式を取るかもしれないが、2つの特定のサーチ手続きによって説明することができる。

【0064】第1の手続きは、「縦続」のサーチとして識別することができる。この手続きにおいて、ターゲットのオブジェクトおよびその他すべてのオブジェクトについて一度に1つの記述子の特徴が比較されて、データの組を連続して狭める。例えば、第1の特徴が用いられるサーチの第1段階において、100個のリストからターゲットに「マッチ」する10個のオブジェクトのリストが得られ、第2のサーチ段階（異なる記述子特徴）において、第1段階で見つかった10個の項目からうちの「マッチ」が得られる、というものである。縦続サーチである。

【0065】第2のサーチ手続きは、重み付けした記述子の特徴の組み合わせを用いて比較を行うことを含む。異なる特徴に付ける特定の重みには、性質の相違および特徴のスケールのために、複雑な考慮をしなければならない。縦続アプローチであれば、より確実なアプローチがなされるので、縦続アプローチが好ましい。

【0066】 $C_{feature}$  の後に  $N_{fr}$ 、 $\Delta_{ms}$ 、 $\Delta_{1c}$  を含むランレングスの特徴の組が続くシーケンスが、縦続サーチにおいて有用な1組の特徴であることがわかってい。マッチングテストの結果は、本発明の記述子によって、ターゲットのオブジェクトと移動が同様のオブジェクトを配置する、ということを示している。第1の段階は、動作特性において非常に異なる候補を削除し、第2の段階は、全体の動作の複雑さは同じであるが動作の強さの分布が異なる候補を削除する。マッチは、意味論的にターゲットのシーケンスから非常に相違することがあるが、意味論的に異なるオブジェクト/フレームのMPEG-2/4をベースにした圧縮の複雑さは同様であり得るので、このことは予想することができる。例えば、コンテンツ船のオブジェクトは、移動の複雑さにおいて  $Activity$  の頭および肩と似ている。その場合、ランレングスの特徴を含む比較の第2段階は通常、意味論的に同様のオブジェクトをより近くするが、表面の感じにおける相違によって船を頭および肩の画像から分ける。このような最終識別が、この場合には行われる。これはプロセスの目的が、一層複雑な方法によって更に処理することができる候補の短いリストを生成することであるからである。本システムは、所望の結果を達成する。

【0067】様々なMPEG-4のオブジェクトに関する本発明による技術を用いた第1および第2の段階のマッチングの具体例を、図3、図4、および図5に示す。

【0068】従って、本発明の記述子の方法は、比較的簡単な装置に適用することができ、次の装置またはプログラムによるより小さなデータの組上へのより高レベルのMPEG-7の記述子の計算を促進する、中間MPEG-7記述子として有用であろう。

【0069】前述の技術を用いる上での更なる段階として、時間的分割マーカを生成してもよく、これらのマーカは、入力ビデオシーケンスと関連して、特定のタイプ

の情景変化を配置して識別する。その後、ビデオシーケンスの特徴を本発明に従って抽出することができる。

【0070】突然の情景変化およびフェードイン/フェードアウトの情景変化の検出段階のそれぞれについての具体的な好適な構成は、以前に出願されている“methods of scene change detection and fade detection for indexing video sequences”という名称の出願において詳細に説明されている。本発明のより一般的な態様から逸脱することなく、情景変化を検出する各段階の詳細を変える機会が本質的にある、ということが理解されるべきである。

【0071】簡単に言えば、情景変化を検出する好適な方法は、以下の各段階を含む。

【0072】1. 情景変化が存在すると疑われるGOPを、DC-画像をベースにしたプロセスを用いることによって、連続したIフレーム上に配置する。

【0073】2. 段階1において選択されたそれぞれのGOPにおいてビット割り当てをベースにした基準を適用して、切点を配置する。

【0074】この技術をMPEG-4圧縮ビデオに適用するには、以下のより詳細な基準が用いられる。

【0075】MPEG-4はオブジェクトをベースにしているので、2つの隣接するフレームにおいて同様に配置されたオブジェクトを表すブロックは、予備限定段階において比較される。それぞれのオブジェクトにおける時間的変化が測定され、1つのフレーム内のすべてのオブジェクトにわたる変化（相違）の重み付けした合計が決定される。この重みは、そのオブジェクトが全フレーム領域に占める割合に関係する。オブジェクトの変化はまた、それぞれのショットまたは情景におけるオブジェクトレベルで手続きを繰り返すことによっても検出される。閾値レベルよりも大きい変化は、情景変化があると疑われるということを示す。

【0076】それぞれのオブジェクトをコード化するのに必要なビット数はかなり変化するので、MPEG-2のデータにおけるような固定した閾値は適当ではない。MPEG-4のデータに固定した閾値を用いると、誤った検出を行う、および/またはオブジェクトレベルにおいて情景変化を検出しない、という結果になる。従って、固定した閾値とスライドするウィンドウとを組み合わせたものを用いて、関係するすべてのビットレートにおいて検出が確実に行われるようにする。

【0077】1つのフレーム速度から他のフレーム速度に変化する場合には、フレーム複製が用いられる場合がある、ということも考慮に入れるべきである。ビット割り当てのビット画像を、一方は移動ベクトルについてのビット割り当てに対応し他方は残りに対応する、2つの画像に分割することによって、フレーム複製を、移動ベクトルをコード化するのに必要なビット数の急激な低下、およびそれに伴う、残りをコード化するのに必要な

ビット数の降下として、検出することができる。両方のパラメータが同時に降下することによって、このデータに、情景変化決定の資格がなくなる。このような工夫を付け加えることによって、フレーム複製を用いた場合の誤った検出の問題が削除される。

【0078】突然の情景変化の検出に用いる具体的な1組の段階は、以下のとおりである。

【0079】1. テストシーケンスSが、フレーム $I_0, I_1, \dots, I_N$ からなるとする。

【0080】2. 1フレーム $I_0, I_1, \dots$ からなるサブシーケンス $S_1$ を形成する。

【0081】3. DC画像をベースにした基準を用いて、 $S_1$ の情景変化検出を実行する。すなわち、それぞれのマクロブロックについて、そのマクロブロックにおけるそれぞれの $8 \times 8$ のブロックについてのDC値の合計の $1/4$ に等しいDC値を決定し、時間的に隣接するマクロブロック同士を比較して、選択した閾値よりも相違の方が大きい場合に潜在的情景変化を決定する。フレーム同士を比較する間、フレームにおいて最も密接した対応するスペースを占めるオブジェクトが比較され、比較されているオブジェクト同士のサイズが異なる場合には、移動推定プロセスを用いることによって、その2つのオブジェクトの間での最良のマッチを得て、相違の基準を計算する。それぞれのオブジェクトOの領域を $A$ 、その相違の基準を $d$ 、フレーム $n$ に対応する相違の基準を $D_n$ とすると、基準全体は、以下のようになる。

【0082】 $D_n = \sum A_i * d_i / \sum A_i$

【0083】4. 前段階から、情景変化を有すると疑われる対 $I_n, I_{n+1}$ を識別する。対応するGOPを、 $G_{test(n)}$ 、 $G_{test(n+1)}$ とする。

【0084】5. それぞれのGOPについて、前段階から、以下のものを用いて情景変化検出を実行する。

(a) マクロブロック当たりのビットの基準を用いた、PフレームとPフレームとの比較。

(b) 前段階において情景変化が検出されていない場合には、テストしたGOPの前のGOPを $GOP_{prev}$ 、テストしたGOPの次のGOPを $GOP_{next}$ とする。 $GOP_{prev}$ の最後のPフレームを取り出し、それを、マクロブロック当たりのビットの基準を用いて、 $GOP_{test}$ のPフレームと比較し、同様に、テストしたGOPの最後のPフレームを取り出し、それを $GOP_{next}$ の最初のPフレームと比較する。情景変化が検出される場合には、対応するIフレームにおいて情景変化を宣言するか、そうでなければ、ビット割り当てをベースにしたアルゴリズムの失敗を宣言する。アルゴリズムの失敗の場合には、DC画像をベースにしたアルゴリズム等の他の技術を用いる。

【0085】6. 上述の情景変化検出をベースにして、Sをショットに分割する。

【0086】7. MPEG-4の場合には、それぞれの

ショット内で同じ手続きを繰り返して、突然のオブジェクトの変化を検出する。

【0087】(ビット割り当てをベースにした基準)それぞれのオブジェクトについて、それぞれのマクロブロックをコード化するのに必要なビット数からなる「ビットレートマトリクス」を構成する。 $R = \{r(i, j)\}$ および $S = \{s(i, j)\}$ を2つのレートマトリクスとすると、RとSとの間のレートの相違は、以下のようになる。

【0088】 $d(R, S) = \sum |r(i, j) - s(i, j)|$

【0089】ただし、 $r(i, j)$ は、オブジェクトの $(i, j)$ 番目のマクロブロックに必要なビット数である。レートマトリクスのシーケンス $R_m, m=1, 2, \dots, N-1$ が与えられれば、次式のように正規化したレートの相違のシーケンス $RP_m, m=1, 2, \dots, N$ を得ることができる。

【0090】 $RP_m = d(R_m, R_{m+1}) / T_m$

【0091】ただし、 $T_m$ は、シーケンス $RP_m$ の $m$ 番目のオブジェクトに必要な合計ビット数である。局所的変化を捕捉するために、スライドするウィンドウを用いてレート相違のシーケンス $RP_m$ を処理する。以下の場合には、 $R_i$ から $R_{i+k}$ までの情景変化を宣言する。

【0092】1. 相違 $RP_i$ が、サイズが $2k-1$ の対称的なスライドするウィンドウ内で最大であり、かつ、

【0093】2. 相違 $RP_i$ がまた、スライドするウィンドウ内で2番目に大きい値の $n$ 倍である。説明する例においては、 $k=3$ および $n=2$ を用いている。これらのパラメータの選択は、経験的データ、フレーム速度、周期的Iフレームの頻度、等によって決まる。速度の相違は、2つのIフレーム/オブジェクトの間、または2つのPフレーム/オブジェクトの間のみで計算することができる、ということに注意されたい。言い換えれば、シーケンス $R_i$ におけるすべてのフレームは、Iフレーム(オブジェクト)かPフレーム(オブジェクト)かのどちらかのはずである。また、MPEG-4のシーケンスにおいては、必ずしもすべてのオブジェクトを内部コード化で同時に「リフレッシュ」する必要はない、ということにも注意されたい。

【0094】(DCシーケンスをベースにした基準)このアプローチにおいては、DC画像(オブジェクト) $X = \{dc(i, j)\}$ のシーケンスを構成する。ただし、 $dc(i, j)$ は、画像(オブジェクト)の $(i, j)$ 番目のマクロブロックのDC値である。DC値を内部コード化フレームまたはオブジェクトから抽出することは、エントロピー復号が必要なのであるが、簡単であるが、予測フレーム/オブジェクトおよびPフレーム/オブジェクトからDC値を抽出することは、それよりも多くの計算およびメモリが必要である。その後、いくつかの可能性のある計量のうちの1つを用いて、前のシ

ーケンスにおけるように、相違のシーケンスを構成する。2つのフレームXとYとの間で、以下のように定義される計量 $d_c(X, Y)$ を用いる。

【0095】 $d_c(X, Y) = \sum (|x_{i,j} - y_{i,j}|)$

【0096】従って、このシーケンスについてシーケンス $d_c(X_i, X_{i+1})$ を構成し、前述したスライドするウィンドウのアプローチを用いて、情景変化を検出することができる。

【0097】(残りのDC成分をコード化するのに必要なビットの変化を用いたフェード検出) 徐々に情景が変化する場合、画像が完全にブラックのフレームからフェードインしているか、完全にブラックのフレームにフェードアウトしているかであるので、画像のそれぞれのブロックがDC修正項を含む、ということが観察されている。このことを頭に置いて、残りのブロックのDC成分のビット割り当てのプロファイルによって、フェード検出のインジケータが提供される、ということがわかっている。本発明の発明者が同時に出願している第2の出願においてより詳細に説明されているフェード検出方法は、一般的に以下の各段階を含む。

【0098】1. 連続したIフレームのDC画像を比較して、情景変化が疑われるものを配置する。このことは、上述の突然の情景変化を検出する方法に必要であることがわかっている。この後、連続するIフレームが非常に異なる信号セグメントにおいてのみ情景変化のサーチを行い、従って信号シーケンス全体を処理することが回避されるので、この段階は、計算を省くのに役立つ。

【0099】2. 情景変化が疑われるものがある領域におけるそれぞれのPフレームについて、負のDC成分を有するブロック数、および正のDC成分を有するブロック数がカウントされる。MPRG-2の場合もMPRG-4の場合も、この段階は、VICパースから容易に行うことができる。これはそれぞれのゼロでないDC成分に、ゼロでない数のビットおよびその成分が正か負かを示すサインビットが割り当てられるからである。ゼロのDC成分は、ランレングスによって示され、従って容易にスキップすることができる。

【0100】3. フレーム数に対する、上で得られた2つの数の特性を決定し、情景変化が疑われるものが、上記段階1に従って配置されている領域を決定する。

【0101】4. 負の遷移の数が著実にゼロでない遷移の総数の60%よりも大きいとそれと等しい場合には、フェードアウトを宣言する。逆に、正の遷移の数が前述の閾値を満たす場合には、フェードインを宣言する。前述の60%の閾値の代わりに、スライドするウィンドウの1バージョンを用いてもよい、ということに注意するべきである。

【0102】この検出体系は、エンコーダが実行する移動のサーチを利用するものである。あるマクロブロックと、前のフレームにおいてそれと最も密接にマッチする

ものとの間のDCの相違のコード化に費やす速度を測定する。

【0103】要約すれば、オブジェクト/フレームにおけるそれぞれのマクロブロックの移動ベクトルおよび残りをコード化するのに必要なビット数をベースにして、記述子を構成する。

【0104】移動ベクトルおよび残りの圧縮複雑マトリクスは、できるだけビットレートから独立させる。

【0105】フレーム/オブジェクトのマクロブロック当たりの移動ベクトルに費やす平均ビット数を計算することができる。この平均は、フレーム/オブジェクトの移動の複雑さの有用な更なる尺度である。MPRG-4で低ビットレートの場合には、この平均によって、静止したオブジェクトと動いているオブジェクトとが容易に識別される。同様に、フレーム/オブジェクトのマクロブロック当たりの残りの平均の圧縮の複雑さを計算する。この数は、オブジェクト/フレームを分類する/マッチさせるのに役立つ。ランレングスを用いて2つのマトリクスを表現する。これは、多くの要素をゼロにさせる低ビットレートにおいて、特に有用である。便利のために、走査順を用いているが、いかなる他の順を用いてもよい。ランレングス構造を用いて、簡単な方法で圧縮の複雑さの空間的分布の形状を表している。ランレングス表現の前に閾値関数が提供され、特定の閾値Tよりも低いすべての要素がゼロにセットされる。閾値は、記述子の一部として供給される。

【0106】これによって、シーケンスのそれぞれのフレームについて、記述子が作り出される。1ショットについての記述子は、まず、上述の情景変化検出技術を用いて、ビデオシーケンスを時間的にショットに分割することによって、決定される。平均付近のフレームのグループ内にあるショットのフレームのフレーム記述子を、そのショットについての記述子として用いる。最初の少数のフレームをスキップして、速度制御を情景の遷移から落ち着かせてもよい。

【0107】マクロブロック当たりの移動ベクトルに費やす平均ビット消費量、およびマクロブロック当たりの平均の圧縮の複雑さは、候補をランク付けするのに役立つマッチング基準の役割をすることができる。そうすると、それぞれの表現におけるランレングス情報を用いて、候補を更にランク付けすることができる。その後、等しく索引付けしたランレングスが比較される、すなわち、フレーム/オブジェクトAの短い、中間の、および長いランレングスのランレングスが、フレーム/オブジェクトBの対応するランレングスと比較される。

【0108】本発明を、異なるプログラム源からの単一のショット同士をマッチさせる点から説明した。しかし前述の技術を、それぞれの集まりが単一のプログラム源に由来するショットの集まりに適用することも、本発明の範囲内にある、ということが理解されるべきである。

例えば、アクション映画は1行にいくつかの移動の複雑さが高いショットを有するが、それよりも静かな映画であれば、1行に時間的空間的複雑さが低いショットを多数有する。これらの特性をログして、マッチング基準を提供するのに用いてもよい。

【0109】また、上述の各段階のシーケンスは、必ずしも互いに時間的にすぐ近接して起こるわけではなく、実際には、時間的に離れているかもしれない、ということにも注意するべきである。特に、記述子を、作り出してプログラムのマテリアルに添付することができ、後で、マッチを見つけるのに用いてもよい。いずれにせよ、このような活動は、本発明の範囲内にあると考えられる。

【0110】従って、本発明を、その好適な実施形態および様々な変形の点から説明したが、当業者であれば、添付の特許請求の範囲に述べる本発明の範囲から逸脱することなく、更なる変形を行うことができる、ということが理解されよう。

【0111】

【発明の効果】本発明は、MPEG-2およびMPEG-4の圧縮ビデオシーケンスからの特徴抽出の方法を提供する。圧縮ビットストリームを検査することによって、ビデオシーケンスの時間的空間的圧縮の複雑さが特徴抽出について評価され、その複雑さを、ビデオシーケ

ンスの時間的空間的特性の記述子として用いる。時間的空間的圧縮の複雑さの尺度は、マッチング基準として用いられ、また、絶対索引付け用に用いることもできる。特徴抽出は、情景変化検出技術と共に行うことができ、この組み合わせは、合理的な正確さを有し、非常に簡単であるという利点を有する。これは、圧縮した形の信号のエントロピー復号をベースにしており、計算的に高価な逆離散のコサイン変換(DCT)を必要としないからである。

【図面の簡単な説明】

【図1】 HDTVの当業者に公知の1組のMPEG-4のオブジェクトを、本発明の仕様を説明する本発明の原理に従って取り出した記述子と共にリストした表を示す図である。

【図2】 MPEG-2のフレームについての時間的空間的複雑さの例を示す図である。

【図3】 第1段階のマッチングの具体例を示す図である。

【図4】 第2段階のマッチングの具体例を示す図である。

【図5】 MPEG-4のオブジェクトに関する本発明による技術を用いたマッチングの具体例を示す図である。

【図3】

ターゲットのオブジェクトと 類似性の降順のオブジェクト	マクロブロック当たりの 平均移動ベクトルビット消費量
Akiyo-頭および肩	1.77
コンテナ船-旗	2.0
沿岸警備隊-モータ発進	2.0
コンテナ船-船	1.16
ニュース-ニュース読者	1.1
Akiyo-フル	0.87

【図1】

MPEG-4圧縮ビデオシーケンス  
 フレームサイズQCIF (176×144)  
 フレーム速度30フレーム/秒  
 圧縮速度1-1.2Mbps

シーケンス	オブジェクト	MB での VOP サイズ	C <sub>avg</sub> mv	C <sub>avg</sub> res	QP	N <sub>0</sub>	最も普通の ランレングスの 範囲	最も普通の ランレングスの 出現の頻度
ニュース	静止背景	11×9	0.2	2.28	3	89	長い	3
ニュース	ダンサーの シーケンスを 有するTVモニタ	6×5	11.43	193.5	3	18	短い	3
ニュース	ニュース読み	11×7	1.1	42.19	3	30	短い	4
ニュース	テキストオーバーレイ	3×2	0.0	0.0	3	6	長い	1
沿岸警備隊	水	11×6	5.48	41.5	3	51	長い	4
沿岸警備隊	モータ発進	9×5	2.00	35.6	3	23	長い	2
沿岸警備隊	小型モータボート および航路	11×1	4.00	63.8	3	8	長い	3
沿岸警備隊	パンする背景	11×4	2.86	51.0	3	36	中間	4
コンテナ船	水	11×8	0.85	5.46	6	58	短い	10
コンテナ船	船	9×4	1.16	46.13	6	15	長い	4
コンテナ船	小型ボート	4×1	0.8	10.4	6	3	長い	1
コンテナ船	前景 (旗ざお)	11×9	0.12	1.79	6	90	長い	4
コンテナ船	静止背景 (空)	11×3	0.4	2.09	6	29	長い	2
コンテナ船	旗	1×1	2	97	6	0	適用なし	0
Akiyo	静止背景	11×9	0.22	2.93	4	89	長い	3
Akiyo	頭および肩	9×8	1.77	35.3	4	37	短い	5

様々なMPEG-4のオブジェクトについての  
 提案される記述子

【図2】

シーケンス	移動に費やされる マクロブロック 当たりの ビット数	マクロブロック当たりの 複雑さの測定
フットボール (720×480)	14	3332
テアリーダー (720×480)	11.3	3882
スプリント (1920×1080)	30	4577
マーチングバンド (1920×1080)	13.8	4317

【図4】

ターゲットの オブジェクトと 類似性の降順の オブジェクト	最も普通の ランレングスの種類	最も普通の ランレングスの 種類の頻度
Akiyoの頭および肩	短い	5
Akiyoのフル	短い	5
コンテナ船-船	短い	5
ニュース-ニュース読者	短い	4
旗	適用なし	



【図5】

ターゲットのオブジェクトと ターゲットのオブジェクトへの 類似性の降順のオブジェクト	マクロブロック当たりの 平均移動ビット消費量
Akiyo-静止背景	0.22
ニュース-静止背景	0.2
コンテナ船-前景(旗ざお)	0.12
コンテナ船-静止背景(空)	0.4
ニュース-テキストオーバーレイ	0.0
コンテナ船-小型ボート	0.8

フロントページの続き

(71)出願人 597067574

201 BROADWAY, CAMBRIDGE,  
MASSACHUSETTS  
02139, U. S. A.

(72)発明者 アジェイ・ディヴァカラン

アメリカ合衆国、ニュージャージー州、ス  
コッチ・ブレインズ、カントリー・クラ  
ブ・ブルバード 47

(72)発明者 ハイファン・スン

アメリカ合衆国、ニュージャージー州、ク  
ランベリー、キングレット・ドライブ・サ  
ウス 61

(72)発明者 伊藤 浩

アメリカ合衆国、ニュージャージー州、フ  
ォート・リー、ロゼット・ストリート  
2416

09 - EP 1022667

19 - JP 200224590

The allocations of bits for motion and for the residual across the frame are used to form a 'descriptor' for a frame for MPEG-2 cases, and an object for MPEG-4 cases. A descriptor is then formed for the video sequence from the frame signatures, which can be used for matching, classification and searching.

DESCRIPTION - The method involves evaluating the spatio-temporal compression complexity of video sequences for feature extraction, by inspecting the compressed bit stream. The complexity is then used as a descriptor of the spatio-temporal characteristics of the video sequence. The spatio-temporal compression complexity measure is used as a matching criterion, and can be used for absolute indexing.